BEAT
The Behavior Expression Animation Toolkit

Justine Cassell
Hannes Vilhjálmsson
Timothy Bickmore
MIT Media Laboratory

SIGGRAPH 2001

BEAT: Text to Embodied Speech

Goal

- Input is a typed script. Output is automatically produced appropriate nonverbal behavior synchronized with speech

Approach

- Analyze the text for certain linguistic features
- Generate nonverbal behavior based on those features, knowledge bases, and research into human conversational behavior
- Compile the behaviors and schedule them to be animated in synchrony with speech

SIGGRAPH 2001

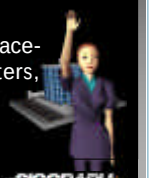
Previous Work:

- Automatic lip synchronization (Waters, 1994)
- Talking heads (Nagao & Takeuchi, 1994)
- Automatic animation of comics from text (Kurlander et al., 1996)
- Behavior scripting for interaction behaviors (Perlin & Goldberg, 1996)
- Expressive qualities of human gestures (Chi et al. 2000)
- Embodied Conversational Agents (Rickel & Johnson, Lester, André & Rist)

SIGGRAPH 2001

Embodied Conversational Agents

- Animated humanoid agents which integrate speech and synchronized facial and gestural behaviors.
- Naturalistic procedural animation of face-to-face conversation — among characters, or between characters and humans.



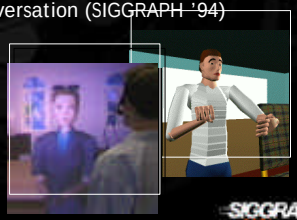
SIGGRAPH 2001

Our Previous ECAs:

Cognitive Representation to Nonverbal Behavior

Examples:

- Animated Conversation (SIGGRAPH '94)
- REA (CHI '99)



SIGGRAPH 2001

Design Goals of BEAT

Support range of users and architectures

Extensibility & modularity for

- Variety of real-time and off-line animation systems
- Event-based or scheduled animation
- TTS or recorded audio
- Addition of new nonverbal behaviors and theories of face-to-face conversation
- Porting to new applications & domains

Authorial control

- Give animators ability to augment and override BEAT's choices

SIGGRAPH 2001

Design Features

XML pipeline architecture

- supports extensibility and modularity
- many extensions can be made in XSLT

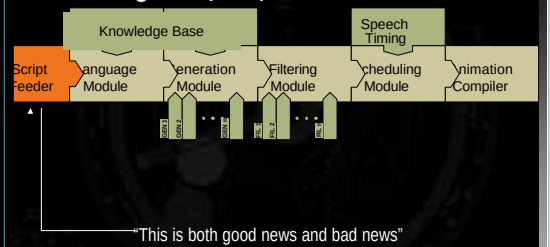
Separation of *generation* and *filtering* of nonverbal behaviors

- provides greater range of possible character behavior and allows multiple generation algorithms to be integrated

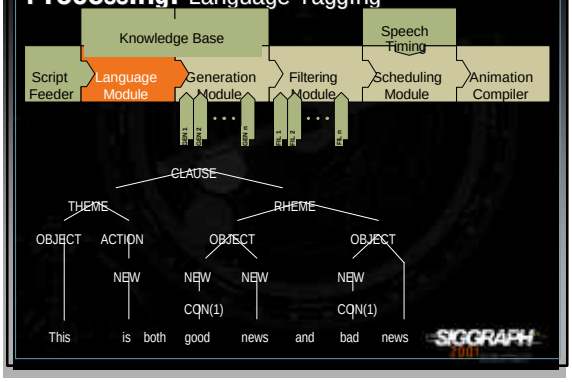
Implemented in Java

- supports portability

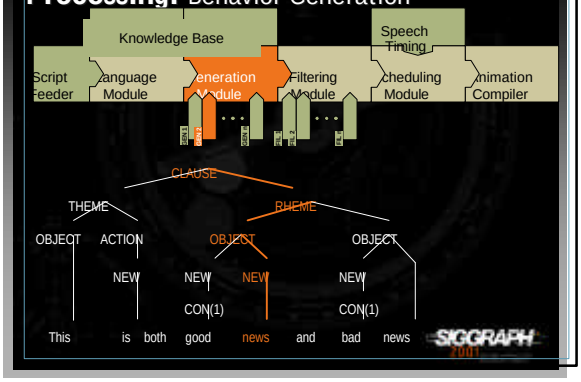
Processing: Script Input



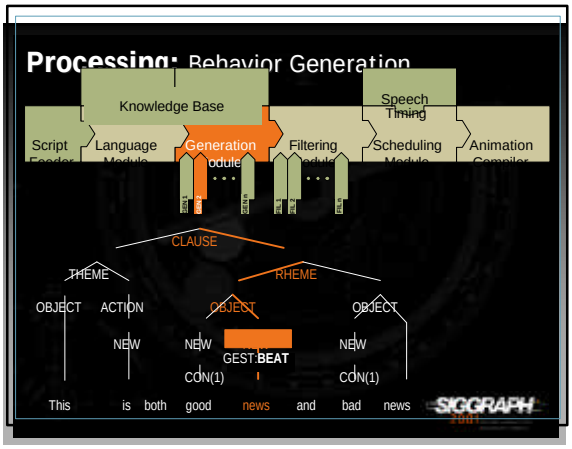
Processing: Language Tagging



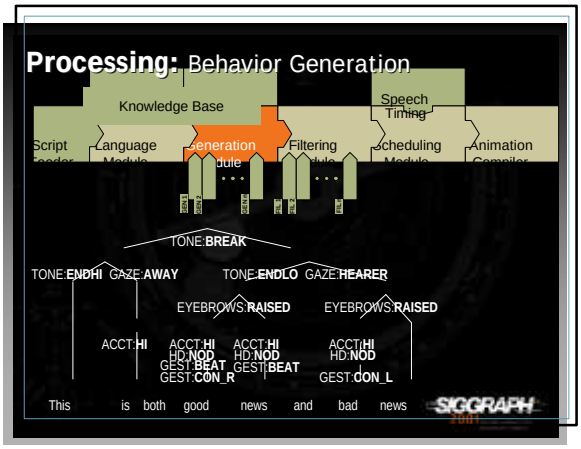
Processing: Behavior Generation

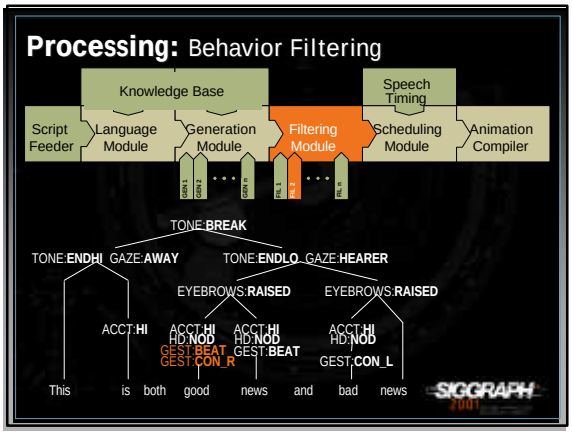


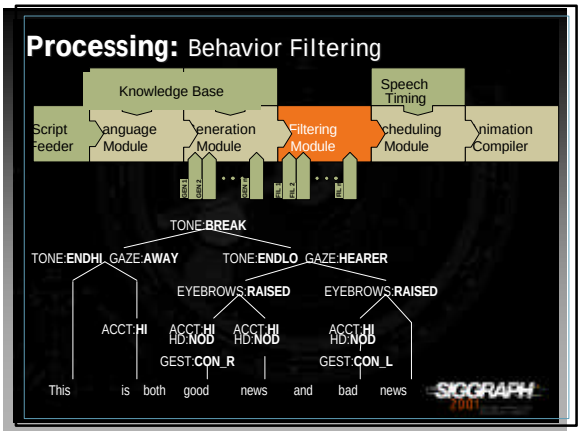
Processing: Behavior Generation

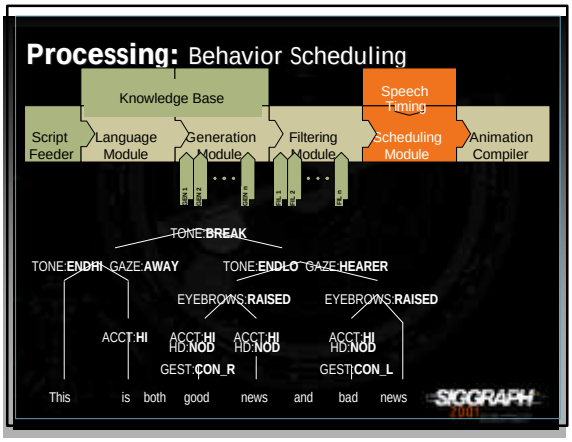


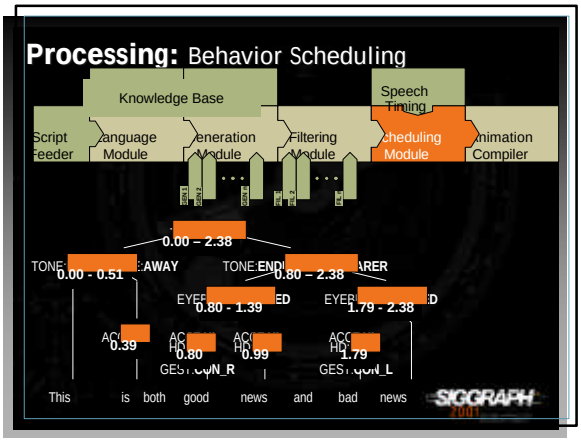
Processing: Behavior Generation

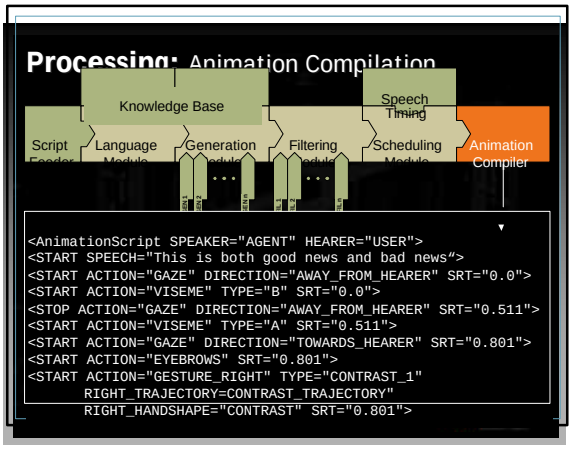


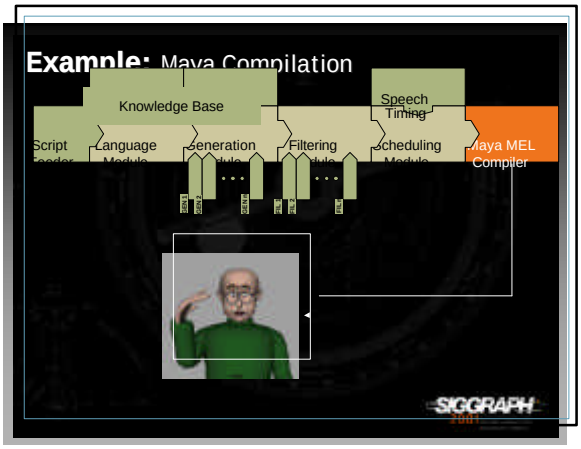

Processing: Behavior Filtering
Knowledge Base
Speech Timing
Script Feeder
Language Module
Generation Module
Filtering Module
Scheduling Module
Animation Compiler
TONE:BREAK
TONE:ENDHI GAZE:AWAY
TONE:ENDLO GAZE:HEARER
EYEBROWS:RAISED
EYEBROWS:RAISED
ACCT:HI
ACCT:HI HD:NOD GEST:BEAT GEST:CON_R
ACCT:HI HD:NOD GEST:BEAT
ACCT:HI HD:NOD GEST:CON_L
This is both good news and bad news
SIGGRAPH 2001


Processing: Behavior Filtering
Knowledge Base
Speech Timing
Filtering Module
TONE:BREAK
TONE:ENDHI GAZE:AWAY
TONE:ENDLO GAZE:HEARER
EYEBROWS:RAISED
EYEBROWS:RAISED
ACCT:HI
ACCT:HI HD:NOD GEST:CON_R
ACCT:HI HD:NOD
ACCT:HI HD:NOD GEST:CON_L
This is both good news and bad news
SIGGRAPH 2001

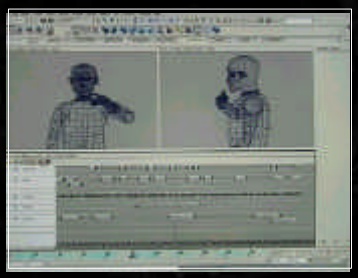

Processing: Behavior Scheduling
Knowledge Base
Speech Timing
Scheduling Module
TONE:BREAK
TONE:ENDHI GAZE:AWAY
TONE:ENDLO GAZE:HEARER
EYEBROWS:RAISED
EYEBROWS:RAISED
ACCT:HI
ACCT:HI HD:NOD GEST:CON_R
ACCT:HI HD:NOD
ACCT:HI HD:NOD GEST:CON_L
This is both good news and bad news
SIGGRAPH 2001


Processing: Behavior Scheduling
Knowledge Base
Speech Timing
Scheduling Module
0.00 – 2.38
0.00 - 0.51
0.80 – 2.38
0.80 - 1.39
1.79 - 2.38
0.39
0.80
0.99
1.79
This is both good news and bad news
SIGGRAPH 2001

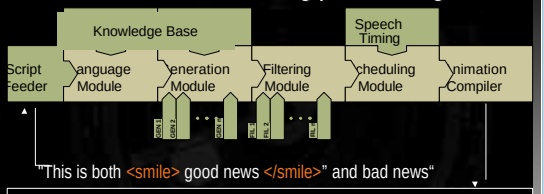

Processing: Animation Compilation
Knowledge Base
Speech Timing
Animation Compiler
<AnimationScript SPEAKER="AGENT" HEARER="USER">
<START SPEECH="This is both good news and bad news”>
<START ACTION="GAZE" DIRECTION="AWAY_FROM_HEARER" SRT="0.0">
<START ACTION="VISEME" TYPE="B" SRT="0.0">
<STOP ACTION="GAZE" DIRECTION="AWAY_FROM_HEARER" SRT="0.511">
<START ACTION="VISEME" TYPE="A" SRT="0.511">
<START ACTION="GAZE" DIRECTION="TOWARDS_HEARER" SRT="0.801">
<START ACTION="EYEBROWS" SRT="0.801">
<START ACTION="GESTURE_RIGHT" TYPE="CONTRAST_1"
RIGHT_TRAJECTORY=CONTRAST_TRAJECTORY"
RIGHT_HANDSHAPE="CONTRAST" SRT="0.801">


Example: Maya Compilation
Knowledge Base
Speech Timing
Maya MEL Compiler
SIGGRAPH 2001

Video:
Real-Time Control of Maya



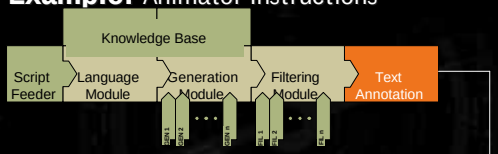
Additional Feature: Tag pass-through



"This is both <smile> good news </smile>" and bad news"

```
<AnimationScript SPEAKER="AGENT" HEARER="USER">
<START SPEECH="This is both good news and bad news">
<START ACTION="GAZE" DIRECTION="AWAY_FROM_HEARER" SRT="0.0">
...
<START ACTION="VISEME" TYPE="B" SRT="0.801">
<START ACTION="SMILE" SRT="0.801">
<START ACTION="GAZE" DIRECTION="TOWARDS_HEARER" SRT="0.801">
<START ACTION="EYEBROWS" SRT="0.901">
...
<STOP ACTION="SMILE" SRT="1.5">
```

Example: Animator Instructions

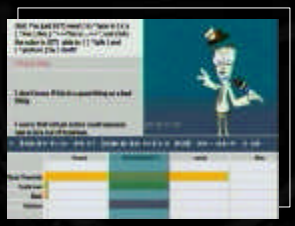


"You just need [1 to *type in] { a [2 *line] like }"

[1] ICONIC - Typing action
[2] BEAT - Emphasis

* Pitch accent
{} Raised eyebrows
.. Gaze away
__ Gaze towards

Video:
Animation from BEAT-annotated text



Future Work

BEAT is a flexible platform for procedural character animation of nonverbal conversational behaviors synchronized with speech

Future work:

- More complete coverage of conversational behavior
- Extending to multiple characters
- Extending to additional animation systems
- Speed

Currently in use by three other research groups

Acknowledgements

MIT

- Yang Gao, Ian Gouldstone and the other members of GNL

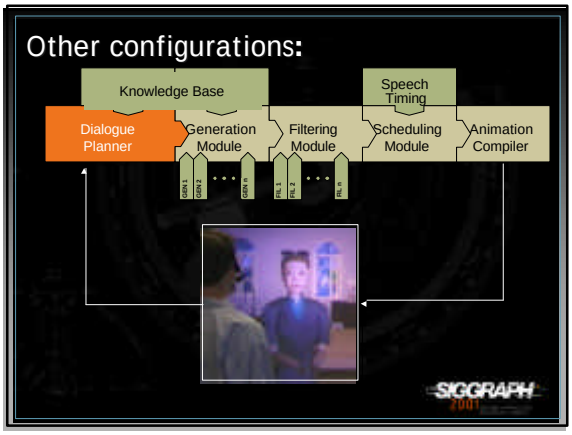
Nearlife

- Dennis Bromley, Geoffrey Beatty, Steve Curcuru, Ryan Kavanaugh

Alias Wavefront

- Jerome Maillot

For more information:
gn.www.media.mit.edu/groups/gn/


Other configurations:
Knowledge Base
Speech Timing
Dialogue Planner
Generation Module
Filtering Module
Scheduling Module
Animation Compiler
SIGGRAPH 2001